ClustalW הכלי

ClustalW– תُستخدم هذه الأداة لمُقارنة تسلسلَيْن أو أكثر، بواسطة طريقة تُسمى التراصف التسلسلي (הַעֲמָדַת רצפים- Sequence Alignment).

ClustalW– تُستخدم هذه الأداة لمُقارنة تسلسلات من الأحماض الأمينيّة أو تسلسلات من النوكلوئتيدات بواسطة طريقة تُسمّى التراصف التسلسلي (הַעֲמָדַת רצפים- Sequence Alignment). مُقارنة تسلسلَيْن – أو أكثر- تُمكِّن من التعرُّف على المناطق وعلى المواضع التي تتشابه فيها التسلسلات، وكذلك على المناطق والمواضع التي تختلف فيها التسلسلات أيضًا.

لعمليّة المُقارنة بين التسلسلات أهميّة كبيرة في البحث البيوإنفورماتي. المناطق المتشابهة في التسلسلات الموجودة في الأنواع المُتشابهة تدلّ عادة على وظيفة و/ أو مبنى مُتشابهين. تُسمّى التسلسلات المُتشابهة أيضًا باسم تسلسلات هومولوجية. فيما يلي عدَّة أمثلة نستخدِم فيها الأداة ClustalW لإجراء تراصف تسلسلات :

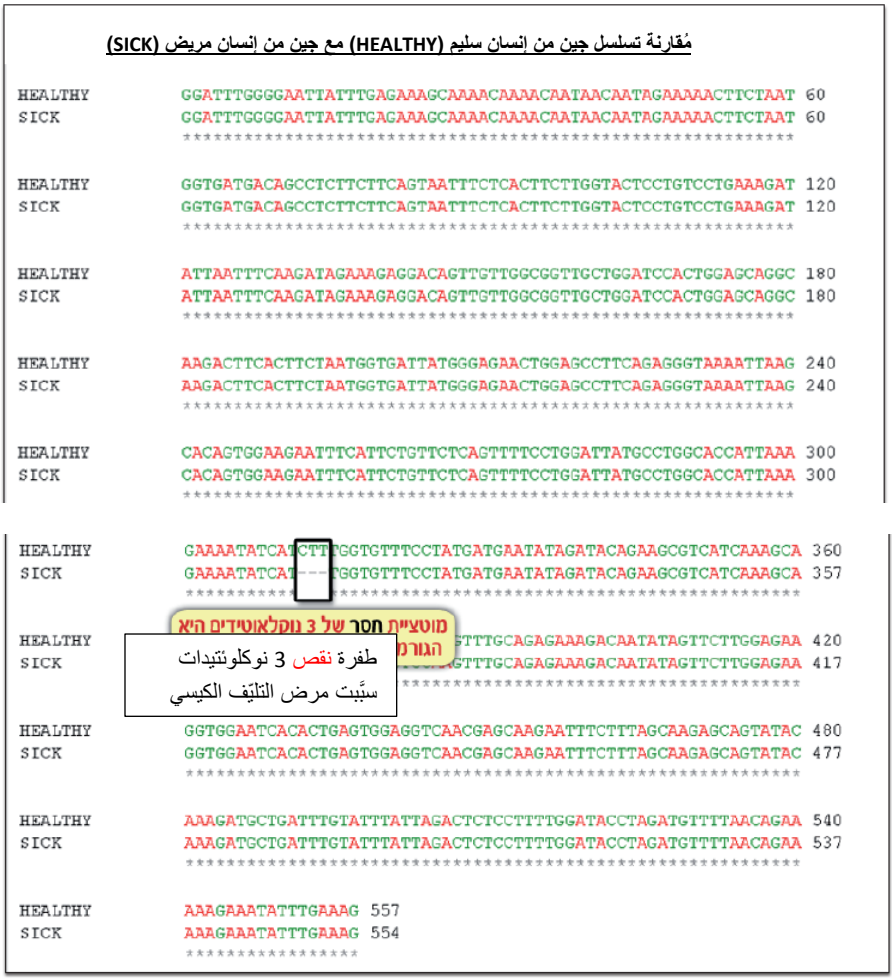
- مُقارنة بين تسلسل نوكلوئتيدات جين سليم وجين طافر تؤدّي إلى الكشف عن مكان الطفرة في التسلسل وكذلك عن نوع هذه الطفرة (طفرة استبدال، طفرة نقص وغيرها). بشكل مُشابه بالإمكان المُقارنة بين تسلسل بروتين سليم وتسلسل بروتين طافر والمُقارنة بينهما.
- مُقارنة بين تسلسلات بروتينيّة مُتشابهة، مثلًا بروتينات تابعة لنفس العائلة البروتينيّة، تعرض المناطق المُتشابهة (المناطق المحفوظة) في هذه البروتينات وأماكنها كما وتعرض المناطق المُختلفة بينها أيضًا. منطقة محفوظة بإمكانها أنْ تُشير إلى مبنى ووظيفة مُتشابهين.
- مُقارنة بين تسلسل RNA رسول (المكوّن من اكسونات فقط) وبين مقطع جين من الكروموسوم (الجين يضمّ اكسونات وانترونات)، تُشير إلى المكان الدقيق للاكسونات في تسلسل هذا الجين.
- مُقارنة تسلسلات (جينات أو بروتينات) مصدرها من أنواع مُختلفة من الكائنات تُشير إلى مدى التقارب والتشابه بين هذه الأنواع. تشابه كبير بين تسلسلات من نوعين من الكائنات يدلّ على تقارب كبير بين النوعين (קרבה אבולוציונית) (للتوسع اقرأ في المُقدمة " אבולוציה" في فُصول المُقدمة).

ClustalW– يُمكن الوصول إلى الأداة بواسطة الرابط:

http://www.ebi.ac.uk/Tools/clustalw2/index.html

اهلا بكم في الجولة الإرشاديّة لأداة ClustalW .ClustalW هو برنامج لمُقارنة التسلسلات، Sequence Alignment.

عندما يكون لدينا تسلسلان أو أكثر فإن الأداة تقوم بوضع التسلسلات الواحد مُقابل الآخر لإيجاد المناطق المُتشابهة والمناطق المُختلفة بين التسلسلات. لمُقارنة التسلسلات أهميّة كبيرة من في البحث البيوإنفورماتي. مثلًا، تكشف المُقارنة بين تسلسل من إنسان سليم وتسلسل من إنسان مريض بمرض وراثي عن سبب هذا المرض. في هذا المثال يُمكن أنْ نرى أنَّ طفرة نقص ثلاثة نوكلوئتيدات أدّت إلى مرض التليّف الكيسي.

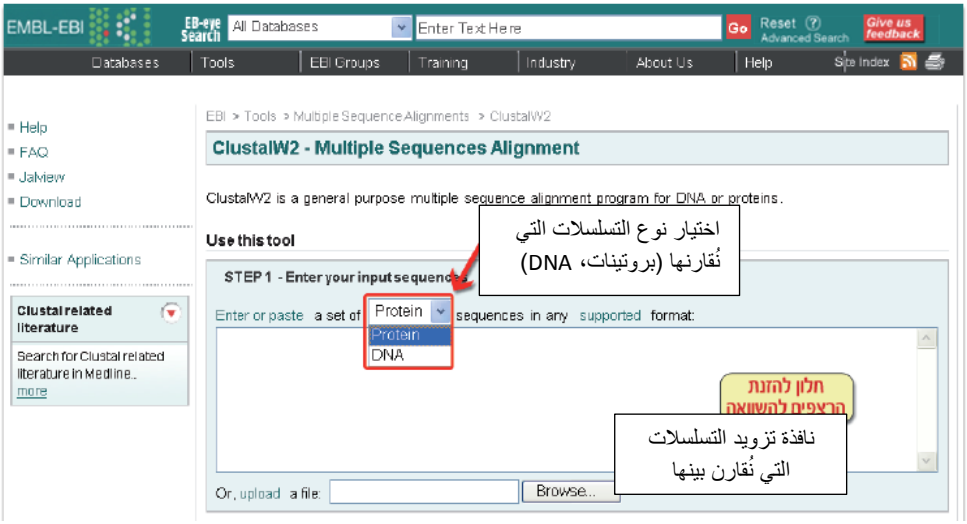


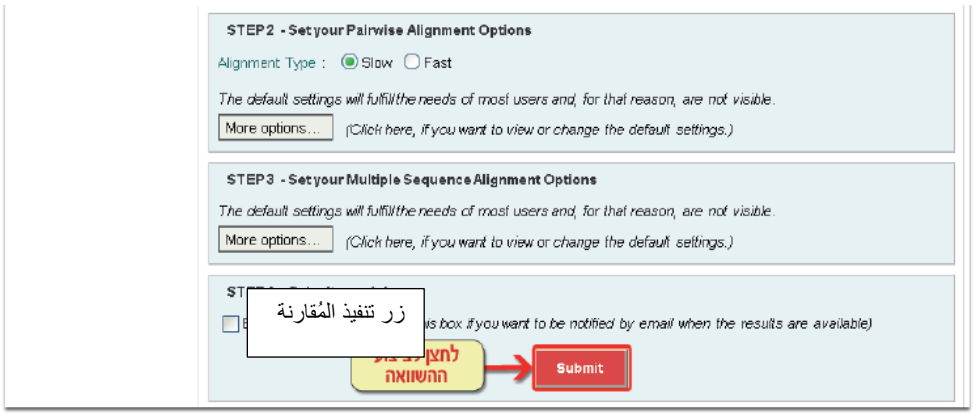
مُقارنة تسلسل جين من إنسان سليم (HEALTHY) مع جين من إنسان مريض (SICK)

```
HEALTHY         GGATTTGGGGAATTATTTGAGAAAGCAAAACAAAACAATAACAATAGAAAAACTTCTAAT 60
SICK            GGATTTGGGGAATTATTTGAGAAAGCAAAACAAAACAATAACAATAGAAAAACTTCTAAT 60
                ************************************************************

HEALTHY         GGTGATGACAGCCTCTTCTTCAGTAATTTCTCACTTCTTGGTACTCCTGTCCTGAAAGAT 120
SICK            GGTGATGACAGCCTCTTCTTCAGTAATTTCTCACTTCTTGGTACTCCTGTCCTGAAAGAT 120
                ************************************************************

HEALTHY         ATTAATTTCAAGATAGAAAGAGGACAGTTGTTGGCGGTTGCTGGATCCACTGGAGCAGGC 180
SICK            ATTAATTTCAAGATAGAAAGAGGACAGTTGTTGGCGGTTGCTGGATCCACTGGAGCAGGC 180
                ************************************************************

HEALTHY         AAGACTTCACTTCTAATGGTGATTATGGGAGAACTGGAGCCTTCAGAGGGTAAAATTAAG 240
SICK            AAGACTTCACTTCTAATGGTGATTATGGGAGAACTGGAGCCTTCAGAGGGTAAAATTAAG 240
                ************************************************************

HEALTHY         CACAGTGGAAGAATTTCATTCTGTTCTCAGTTTTCCTGGATTATGCCTGGCACCATTAAA 300
SICK            CACAGTGGAAGAATTTCATTCTGTTCTCAGTTTTCCTGGATTATGCCTGGCACCATTAAA 300
                ************************************************************

HEALTHY         GAAAATATCATCTTTGGTGTTTCCTATGATGAATATAGATACAGAAGCGTCATCAAAGCA 360
SICK            GAAAATATCAT---TGGTGTTTCCTATGATGAATATAGATACAGAAGCGTCATCAAAGCA 357
                ***********************************************************

HEALTHY         [illegible]GTTTGCAGAGAAAGACAATATAGTTCTTGGAGAA 420
SICK            [illegible]AGTTTGCAGAGAAAGACAATATAGTTCTTGGAGAA 417
                ******************************************

HEALTHY         GGTGGAATCACACTGAGTGGAGGTCAACGAGCAAGAATTTCTTTAGCAAGAGCAGTATAC 480
SICK            GGTGGAATCACACTGAGTGGAGGTCAACGAGCAAGAATTTCTTTAGCAAGAGCAGTATAC 477
                ************************************************************

HEALTHY         AAAGATGCTGATTTGTATTTATTAGACTCTCCTTTTGGATACCTAGATGTTTTAACAGAA 540
SICK            AAAGATGCTGATTTGTATTTATTAGACTCTCCTTTTGGATACCTAGATGTTTTAACAGAA 537
                ************************************************************

HEALTHY         AAAGAAATATTTGAAAG 557
SICK            AAAGAAATATTTGAAAG 554
                *****************
```

واجهة الأداة ClustalW

هذه هي واجهة الأداة ClustalW. في البداية يجب أنْ نُشير إلى نوع التسلسلات التي نقوم بمُقارنتها (تسلسل من الأحماض النووية أو تسلسل من الأحماض الأمينيّة) عن طريق اختيار protein أو DNA في النافذة المُناسبة. نقوم بتزويد نافذة تزويد التسلسلات الموجودة في واجهة الأداة بالتسلسلات التي نرغب بمُقارنتها، يجب أنْ تكون هذه التسلسلات بصيغة FASTA، في هذه الصيغة يبدأ التسلسل بالرمز ">" ويليه اسم التسلسل، بعد ذلك يُعرض تسلسل البروتين أو الجين على شكل تسلسل من الحروف، بحيث أنَّ كل حرف يُشير إلى حامض أميني (في تسلسل البروتين) أو إلى نوكلوئتيد (في تسلسل الجين). بعد تزويد التسلسل الأول، نزوّد النافذة بالتسلسل الثاني بنفس الطريقة، وهكذا حتى نزوِّد الأداة بجميع التسلسلات التي نرغب بالمُقارنة بينها.

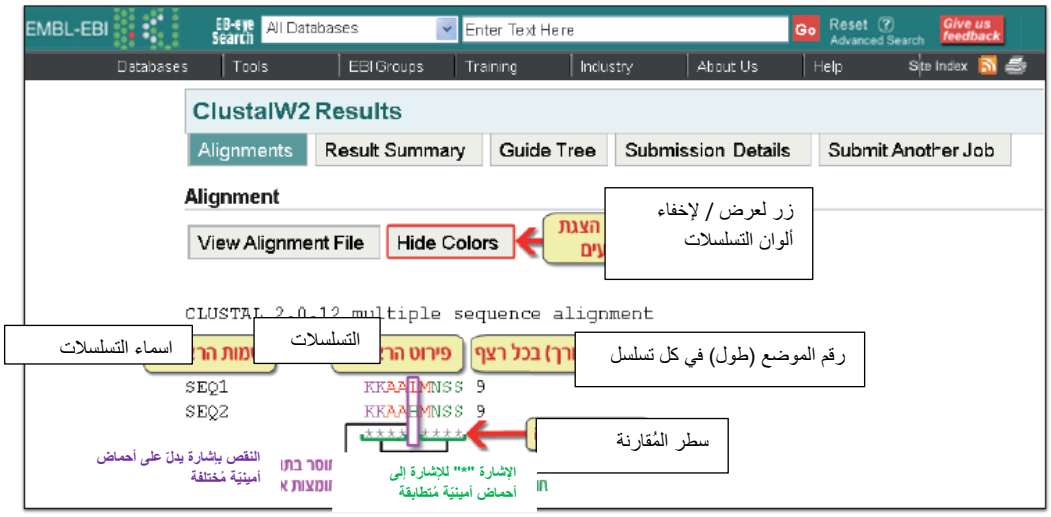




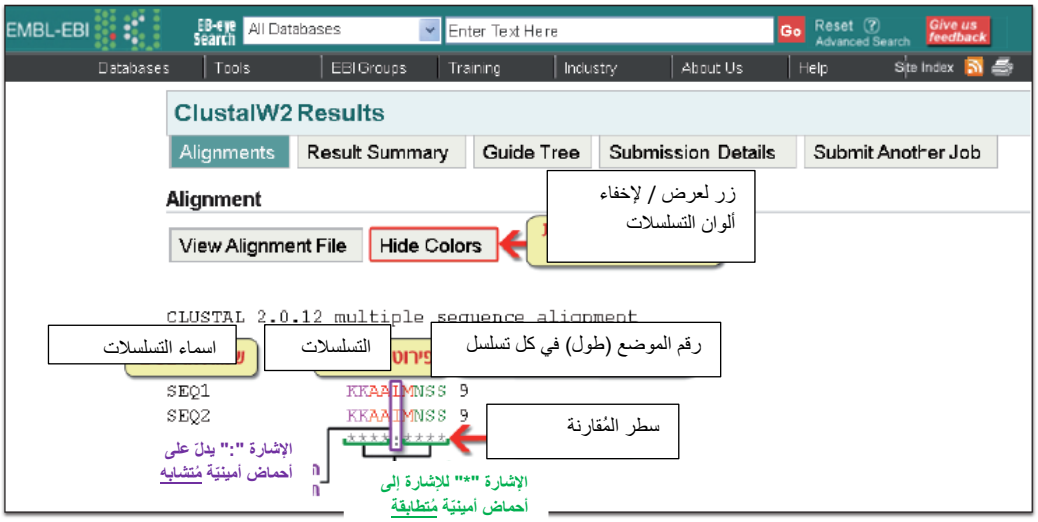
مُقارنة بين تسلسلين

في العرض التوضيحي الأول سنقوم بمُقارنة تسلسلَيْن قصيرَيْن ومُتشابهَيْن جدًّا من الأحماض الأمينيّة. نزوّد الأداة بالتسلسلات التي نرغب بمُقارنتها، نضغط على الزر Submit وننتظر. عملية المُقارنة يُمكن أنْ تستمرّ عدّة ثواني.

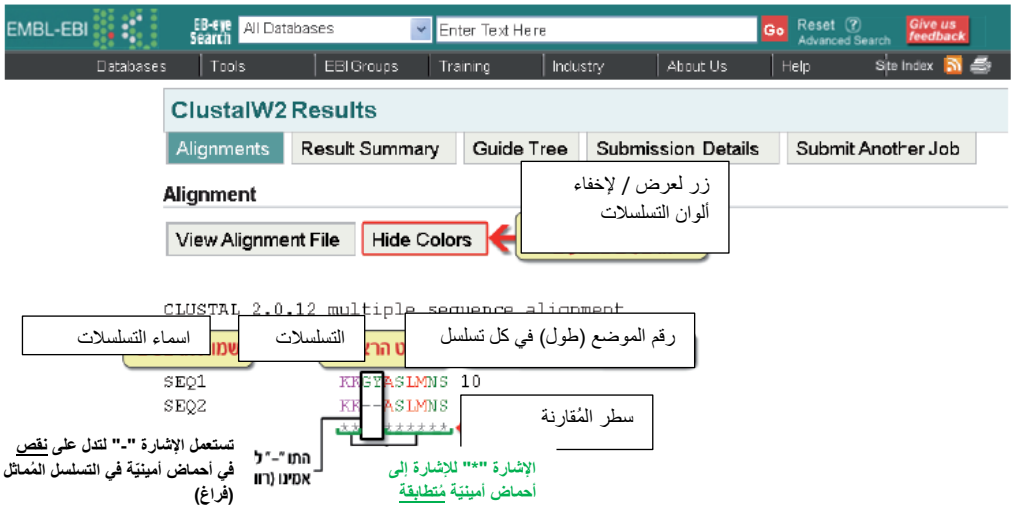
تضمّ صفحة النتيجة التي نحصل عليها تفاصيل كثيرة، لكننا سنُركِّز في بطاقة الاختيار Alignment التي تصف المُقارنة بين التسلسلين. في السطرين الأول والثاني، يتمّ ترتيب التسلسلين الواحد فوق الآخر، بحيث يُعرض اسم التسلسل في بداية كل سطر. عندما نضغط على الزر Show Colors تقوم الأداة بتلوين التسلسلات بحسب صفاتها الكيميائيّة. في السطر الثالث تتمّ الإشارة بالرمز "*" تحت المواضع التي تتطابق فيها الأحماض الأمينيّة في التسلسلين. لا تتمّ الإشارة بتاتًا تحت المواضع التي تختلف فيها الأحماض الأمينيّة بين التسلسلين. من خلال مُقارنة التسلسلات يُمكن أنْ نرى أنَّ الحامض الأميني الخامس فقط يضمّ أحماض أمينيّة مُختلفة: الحامض الأميني ليوسين(L) في التسلسل الأول والحامض الأميني هستيدين (H) في التسلسل الثاني. يختلف هذان الحامضان عن بعضهما بصفاتهما الكيميائيّة والفيزيائيّة بشكل كبير جدًّا، لذلك لا يحتوي السطر الأخير في هذا الموضع على أي إشارة.



نتمعّن في مُقارنة إضافيّة. نُزوِّد النافذة بالتسلسلات ونضغط على Submit.
في هذه المرّة، تُشير نتيجة المُقارنة تحت الموضع الخامس بالإشارة ":"، هذه الإشارة تدلّ على أنّهَ بالرغم من أنَّ الأحماض الأمينيّة الموجودة في الموضع الخامس في التسلسلين مُختلفة، إلّا أنَّ لها صفات كيميائيّة وفيزيائيّة مُتشابهة جدًّا. في هذا المثال الحامض الأميني ليوسين (L) والحامض الأميني ايزوليوسين (I) هي أحماض أمينيّة هيدروفوبيّة.



الأن ننظر إلى مُقارنة تسلسلات ذات أطوال مُختلفة.
نتيجة المُقارنة تدلّ على أنَّه في الموضع الثالث والرابع للتسلسل الأول لا يوجد أحماض أمينية تُقابلها في التسلسل الثاني. هذه الإشارة تُسمّى باللغة الإنجليزية GAP، ويُشار لها في السطر الثالث بإشارة عارضة "-". معنى هذا أنَّ التسلسل الأول أطول من التسلسل الثاني بحامضين أمينيّن.



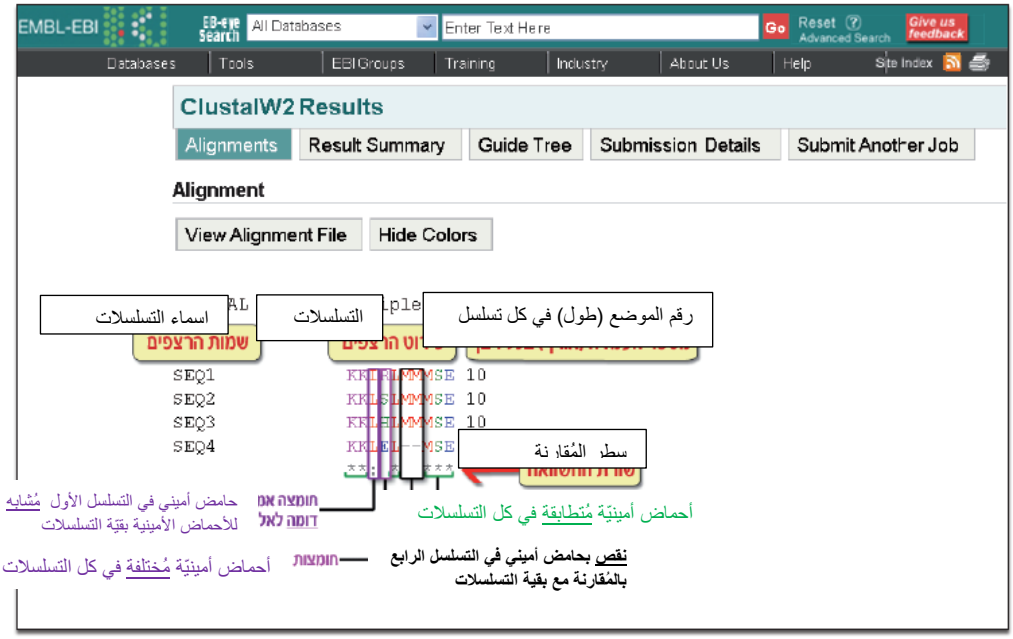
الإشارة		المعنى	
الرمز	الاسم	المُقارنة بين تسلسل أحماض أمينيّة	المُقارنة بين تسلسل نوكلوئتيدات
"*"	نجمة	أحماض أمينيّة مُتطابقة في التسلسلين	نوكلوئتيدات مُتطابقة في التسلسلين
":"	نقطتان	أحماض أمينيّة مُختلفة، لكنها ذات صفات كيميائيّة وفيزيائيّة مُتشابهة جدًّا	
"."	نقطة	أحماض أمينيّة مُختلفة، لكنها ذات صفات كيميائيّة وفيزيائيّة مُتشابهة قليلًا	
"-"	عارضة	نقص حامض أميني في أحد التسلسلين	نقص نوكلوئتيد في أحد التسلسلين
" "	فراغ	أحماض أمينيّة مُختلفة في التسلسلين	نوكلوئتيدات مُختلفة في التسلسلين

المُقارنة بين عدَّة تسلسلات

الأداة ClustalW لا تُلزم المُستخدِم بالمُقارنة بين زوج من التسلسلات فقط. عمليًّا، بالإمكان المُقارنة بين عدد كبير من التسلسلات .

من الصعب التمييز بين المناطق المُتشابهة والمناطق المُختلفة عند مُقارنة عدد كبير من التسلسلات. لذلك نستعين بالأداة ClustalW .

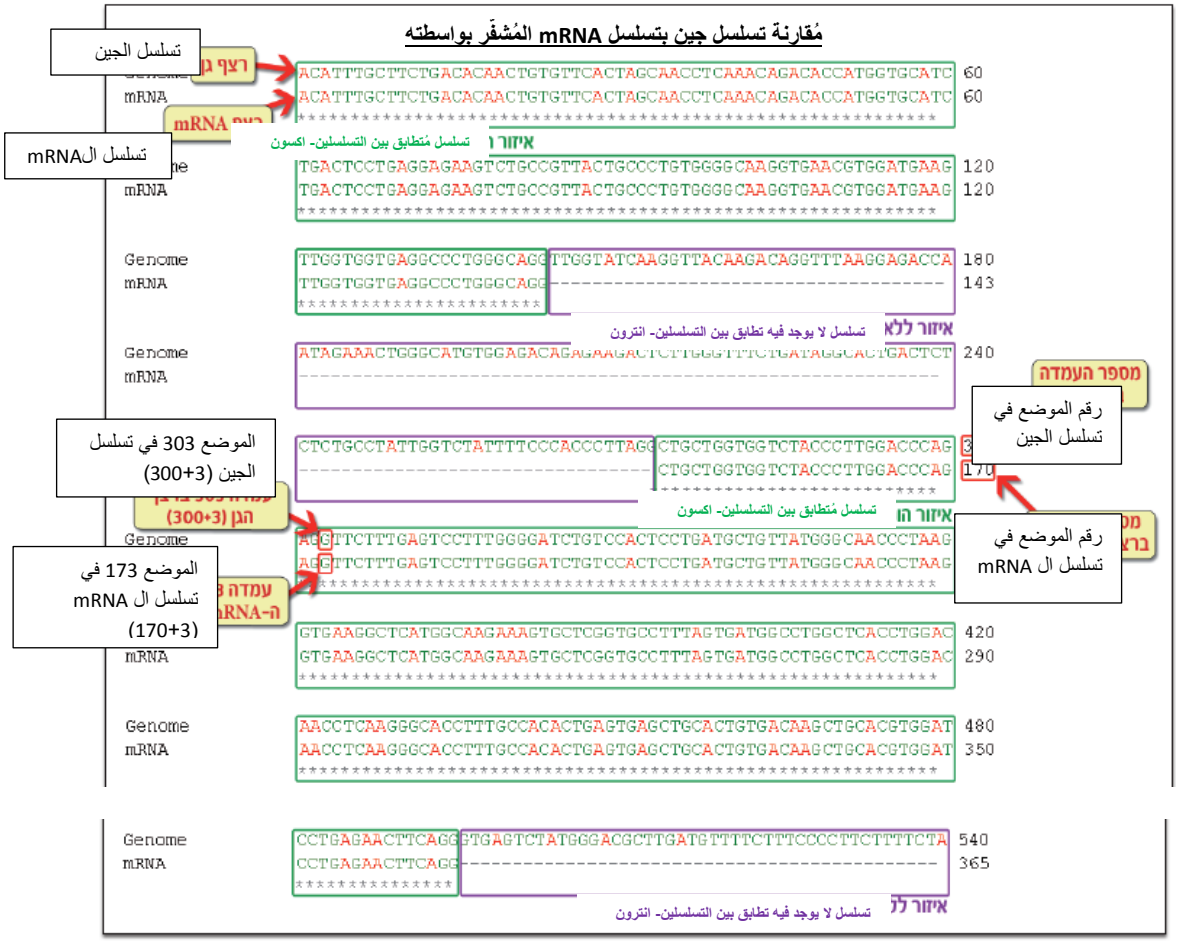
نستطيع أنْ نرى في صفحة النتيجة، تراصف التسلسلات والتمييز بسهولة بين المواضع التي تتواجد فيها أحماض أمينيّة مُتطابقة، مُتشابهة أو مُختلفة في التسلسلات الأربعة المعروضة.



الأمثلة التي عُرضت حتى الآن ضمّت تسلسلات قصيرة جدًّا، عادة تكون تسلسلات البروتينات وتسلسلات النوكلوئتيدات مكوَّنة من تسلسلات أطول بكثير من التسلسلات التي عُرضت في الأمثلة السابقة. مُقارنة مثل هذه التسلسلات وتحليل نتائج المُقارنة بدون استخدام برنامج محوسب هي تقريبًا أمر غير مُمكن ويحتاج إلى وقت طويل. بمُساعدة الأداة ClustalW بإمكاننا مُقارنة عدد كبير من التسلسلات بسهولة وسرعة كبيرتين.

أرقام المواضع في التسلسلات

في مُقارنة التسلسلات تظهر أرقام مواضع في نهاية كل سطر من أسطر المُقارنة، ويُعبر هذا الرقم عن الموضع الأخير في السطر، في السطر الأول للمُقارنة يظهر الرقم 60 بجانب كل واحد من التسلسلين. انتبه لا تكون أرقام المواضع مُتطابقة دائمًا في التسلسلين. مثلًا، في سطر المُقارنة الخامس، رقم الموضع الأخير في تسلسل الجين هو 300 بينما في تسلسل ال RNA رسول الرقم هو 170. انتبهوا: لحساب رقم موضع مُعين داخل التسلسل، يجب أنْ نستعين برقم الموضع الأكثر قُربًا له.



تلخيص

ClustalW هو برنامج لمُقارنة تسلسلات، Sequence Alignment. يُمكن أنْ نُقارن بواسطته تسلسلات من النوكلوئتيدات أو تسلسلات من الأحماض الأمينيّة. يُساعد برنامج ClustalW على وضع التسلسلات مُقابل بعضها ومُقارنتها، حتى نستطيع أنْ نُميِّز بسهولة مواضع الأحماض الأمينيّة المُتطابقة، المُتشابهة أو المُختلفة في التسلسلات. مناطق مُتشابهة (هومولوجيّة) في التسلسلات تدلّ على تشابه في مبنى البروتينات وبالتالي في وظائفها.

أنتم مدعوون لاختبار الأداة والتمرُّن على استعمالها، نتمنّى أنْ تستمتعوا بوقتكم. بالنجاح!